Method of Auto-configuration for Corporate Proxies

Andrés Abelardo Villarroel-Acosta¹, Carlos Enrique Montenegro-Marín², Paulo Alonso Gaona-García³, Yuri Vanessa Nieto-Acevedo⁴

¹ Systems Engineering. Servers Administrator, Universidad Distrital Francisco José de Caldas. Bogotá, Colombia ² PhD in Computer Science. Researcher and teacher, Universidad Distrital Francisco José de Caldas. Bogotá, Colombia. Email: cemontenegrom@udistrital.edu.co

³ PhD in Computer Science. Researcher, Universidad Distrital Francisco José de Caldas. Bogotá, Colombia ⁴ M.Sc. in information and communication sciences. Teacher, Corporación Unificada Nacional (CUN), Bogotá, Colombia

Received date: August 18, 2016 Accepted date: December 6, 2016

How to cite this article: A. A. Villaroel-Acosta, C. E. Montenegro-Marín, P. A. Gaona-García y Y. V. Nieto-Acevedo, "Method of auto-configuration for corporate proxies", *Ingeniería Solidaria*, vol. 13, n.° 21, pp. 9-18, January 2017. doi: http://dx.doi. org/10.16925/in.v13i21.1723

Abstract. *Introduction:* The proxy servers offer many advantages in business and academia. One of their main uses is to protect both the network and its users, offering a secure and fast connection to multiple users who require a service such as Internet. However, manual proxy settings require time and expertise by the end user, which leads to the network being less user friendly. The article was written in 2016 in the faculty of Engineering of the Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. *Methology*: The paper has been elaborated at Colombia using a Cascade Model. *Results*: The article presents an introduction to proxy auto configuration and proxy servers, as well as a detailed explanation of the configuration methods. *Conclusions*: The study demonstrates methods and times that can lead to the reduction in standard connection time through manual configuration vs. automatic proxy configuration. This was implemented as a real case study in a company.

Keywords: dynamic host configuration protocol (DHCP), proxy auto-configuration (PAC), standard time, web proxy auto-discovery protocol (WPAD).



Método de autoconfiguración para proxies corporativos

Resumen. *Introducción*: los servidores proxy ofrecen muchas ventajas en los negocios y la academia. Uno de sus usos principales es proteger, tanto a la red como a sus usuarios, ofreciendo una conexión segura y rápida a múltiples usuarios que requieren un servicio como el Internet. Sin embargo, los ajustes manuales de los sevidores proxy requieren tiempo y experticia por parte del usuario final, lo que hace que la red sea menos fácil de usar. Este artículo fue escrito en el 2016 en la Facultad de Ingeniería de la Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. *Metodología*: el trabajo se elaboró en Colombia utilizando un modelo en cascada. *Resultados*: el artículo presenta una introducción a la configuración automática de proxy y servidores proxy, así como una explicación detallada de los métodos de configuración. *Conclusiones*: el estudio expone métodos y tiempos que pueden conducir a la reducción del tiempo de conexión estándar, comparando la configuración manual con la configuración automática de proxy. Los hallazgos fueron implementados como estudio de caso real en una empresa.

Palabras clave: protocolo de configuración de host dinámico (DHCP), autoconfiguración de proxy (PAC), tiempo estándar, protocolo de descubrimiento automático de proxy web (WPAD).

Método de autoconfiguração para *proxies* corporativos

Resumo. *Introdução*: os servidores *proxies* oferecem muitas vantagens nos negócios e na academia. Um de seus usos principais é proteger tanto a rede quanto seus usuários e oferecer uma conexão segura e rápida para múltiplos usuários que requerem um serviço como a internet. Contudo, os ajustes manuais dos servidores *proxies* requerem tempo e destreza por parte do usuário final, o que faz com que a rede seja menos fácil de usar. Este artigo foi escrito em 2016 na Faculdade de Engenharia da Universidad Distrital Francisco José de Caldas, Bogotá, Colômbia. *Metodologia*: o trabalho foi elaborado na Colômbia utilizando um modelo em cascata. *Resultados*: este artigo apresenta uma introdução à configuração automática de proxy e servidores *proxies*, bem como uma explicação detalhada dos métodos de configuração. *Conclusões*: o estudo expõe métodos e tempos que podem conduzir à redução do tempo de conexão padrão, comparando a configuração manual com a automática de *proxy*. Os achados foram implantados como estudo de caso real numa empresa.

Palavras-chave: protocolo de configuração de *host* dinâmico (DHCP), autoconfiguracao de proxy (PAC), tempo padrão, protocolo de descobrimento automático de *proxy web* (WPAD).



1. Introduction

A proxy server is an application that provides access between two networks, typically a private internal computer network (intranet) and an external computer network such as the Internet [1]. The primary reason for developing the proxy server is to allow access to external sources of computer facilities inside a firewall —protected networks or those that are otherwise directly inaccessible [2]. Proxy servers require each client (such as a web browser program) to be configured so as to recognize and use the proxy servers. The client especially needs to know how to communicate with the proxy, how to format requests to identify the remote servers, and so forth [1]; this manual configuration reduces the connection time to the services offered on the network.

A proxy application offers functionality: If the proxy servers store the contents of the answers, it can improve response times and reduce bandwidth to repeat or allow access restriction to individual client computers; it can detect and block malicious content, perform dynamic recompression and reduce data flow transcription content to skip unsupported contents. It can also increase security by using a proxy server for logging and can be used for anonymous access to a target server [3]. Schools and offices use proxy servers to avoid users from contacting unfortunate websites, and also to examine and prevent them from upsetting customers [4]. It is for this reason that companies use it mostly to improve the security of their own internal network, without this affecting the collateral benefits its implementation might have.

The Ministry of Information and Communications Technologies (Mintic), as head of the technological sector in Colombia and policy leader in the field of information and communication technologies, developed the sectoral and institutional strategic "Plan Vive Digital" (Plan for Digital Living). This is the result of a participatory process with the intervention of entities that are part of the administrative and academic sectors, and that are represented by several universities and relevant members of the general public. One of the major achievements of this plan has been to increase from 2,000,000 internet connections in 2012 to approximately 8,800,000 million in 2014, with forecasts and plans to double that figure for 2017. As part of this development plan and appropriation of information technology in Colombia, a requirement was established: to provide Internet access to students

and teachers in an academic context, ensuring better use. That is why this study develops, implements and analyzes the achievements of the inclusion of a proxy auto-configuration (PAC) in various companies, such as Colombian educational institutions.

This article is structured as follows: in the second section titled "Proxy servers", a theoretical review of the proxy servers is carried out. The third section called "Technologies of implementation", presents the technologies necessary to develop the proposed auto-configurable proxy, analyzing its various methods, providing examples to facilitate understanding. The fourth section with the title "Case study and results" is dedicated to the implementation of the project, followed by its application in a case study, with evidence and results thereof. Finally, conclusions are presented as well as guidelines for future work.

2. Proxy servers

2.1. Web proxy server

A proxy server is usually a computer system -a combination of hardware platforms and software applications- which serves as an intermediary in the network communication between the parties [2]. The client-server systems provide an intermediary in the communication between client (usually sending request) and server (sending the response), as shown in Figure 1. Proxy servers are computers that are positioned between client and server. The client-server is a correlation among two programs, a customer plan that formulates a service demand to a server plan. The basic principle of operation of the proxy server is to receive client requests; these requirements are then analyzed and sent to the target servers, and then the answers pass to the original client in original or modified form.



Figure 1. Proxy server communication Source: [4]

Forward proxies are those in which the client server names the goal server to connect to. Forward proxies are capable of repossessing from a wide variety of sources. The provisos "forward proxy", as well as "forwarding proxy", are common explanations of the performance and thus uncertain [4]. This type of proxy is most frequently used when there is a corporate intranet with users connected to a local area network (LAN). It can also work with a firewall, to provide a security barrier between an internal network and the Internet. The proxy server operates on the 7th ISO/OSI (Open Systems Interconnection model) layer model, to analyze incoming requests [5-6] with the same application protocol, such as service clients. Therefore, this proxy is also known as an application proxy.

As mentioned above, educational institutions and offices use proxy servers to avoid users from contacting undesirable websites. In order to achieve this primary goal, there are two key issues that should be considered if an educational institution or organization uses proxy servers:

- a. Configuring proxy port settings: Corporate administrators and Internet service providers (ISPS) can preset proxy server settings. Once the browser is open, in the Tools menu, Internet options, establish LAN connection setting. The proxy parameters are defined in the "advanced options". Enter the address for the IP/name and proxy server. Next, select use the same proxy server for all protocols". Check the box in the browser or the "use the same proxy server for all addresses", then check the box in the customization wizard that makes all the other entries unavailable and copies the proxy information in the HTTP setting into the other protocol settings. Selecting the check box also hides the information in the socks settings. This configuration is carried out on each client or web browser that needs to use the proxy server. The secure setting is for HTTPS requests based on Secure Sockets Layer (SSL) technology.
- b. Configuring proxy bypass lists: Some networks need to bypass the proxy. The most common reason to bypass the proxy is for local (intranet) addresses. Generally, these addresses do not contain periods in them. By selecting the "bypass proxy server for local (intranet) addresses" check box, all addresses without a period will bypass the proxy and be resolved directly. A proxy

bypass entry may begin with a protocol type: http://; https://; ftp://; or gopher://. If a protocol type is used, the exception entry applies only to requests for that protocol.

If no protocol is specified, any request using the address will be bypassed. If a protocol is specified, requests with the address will be bypassed only if they fit the indicated protocol type. As with the protocol type, address entries are not case sensitive. If a port number is given, the request is processed only if all previous requirements are met and the request uses the specified port number.

2.2. Automatic configuration and automatic proxy

This project aims to eliminate the procedure by which the client must manually configure the proxy in the browser when he or she wants to access a network with this feature; it is therefore important to understand what Proxy Auto-Configuration (PAC) and Web Proxy Auto-Discovery Protocol (WPAD) projects are PAC files written in JavaScript and consist of a set of rules that will allow a browser to determine whether network traffic is accessed directly online or if a proxy server must be used. PAC files provide flexibility and redundancy in a manner which isn't currently possible when configuring an explicit proxy. There are many benefits to using .pac files: they are supported by all major operating systems and browsers, they allow to automatically route traffic correctly, regardless of whether the user enters a domain or IP address and automatic proxy failover [7].

WPAD is a technology which aids a web browser to automatically detect the location of a PAC file using Domain Name Server (DNS) or Dynamic Host Configuration Protocol (DHCP). A browser that supports both DHCP and DNS will first attempt to locate a PAC file using DHCP, and should a DHCP configuration not exist, fail-over to DNS WPAD will occur. If neither are configured, a browser will fail open [7].

a. *Method for WPAD DNS*: The most common method used for automatic detection of PAC files is through the use of DNS; this takes advantage of the name of the user network and DNS settings to discover if the network requires a proxy setting. To use this method you must have a PAC file, a web server and a (hostname) DNS to access locally; the latter must point to the web server. Figure 2 shows traffic information of the WPAD DNS method.



Figure 2. Information traffic for WPAD via DNS Source: Compiled by the authors

Figure 3 shows an example to better understand the WPAD-DNS method. The name of a user network computer is written in: laptop01.us.division.company.com, a PAC file called wpad.dat, a web server stored in wpad.company.com, in which one may see the pac file (Figure 2).

Upon connecting the user computer to the network, the wPAD DNS will replace the computer name (laptop01) with wPAD to then place /wapd. dat as a suffix of the original address. For example, http://wpad.us.division.company.com/wpad.dat; at this point the browser will try to download the PAC file with the name wpad.dar of the previously specified address.

Should the browser not find the file in wpad. us.division.company.com, it will proceed to find the file node sub-domain that is in the hierarchy, (which in the example is wpad.division.company. com). This is the procedure followed until arriving at the lowest node, which will have the PAC file. Finally, the PAC file will be located in wpad.company.com. As mentioned above, this is the address of the web server and that is why it is found here. Figure 3 shows the procedure for performing WPAD by DNs to resolve the problem. This first method is the one most commonly supported by the majority of browsers and operating systems.

b. *WPAD via DHCP method*: In this case, a PAC file is detected to take advantage of the DHCP infrastructure. In order for this to work, it is necessary that in its configuration, the Dynamic Host Configuration Protocol (DHCP) server has the option to store the PAC files so that the browser



Figure 3. WPAD via DNS Source: Compiled by the authors based on [7]

can perform queries. Once the browser finds that this option is enabled, the application will carry out the PAC file. Figure 4 represents traffic information of the WPAD via DHCP method.

Like the previous method, this also has some prerequisites, such as: A PAC file, a web server, a DHCP server and, as already mentioned, that each of the computers of the users who are applying have the proper network settings for a DHCP server.



Figure 4. Information traffic in the WPAD via DHCP Source: Compiled by the authors

An example is shown to facilitate the understanding of the operation of DHCP-WPAD. The network name of a user's computer is used: laptop01.us.division.company.com. Once the name is obtained, the browser issues a DHCP-INFORM, in which the DHCP server provides a list of options and configurations to carry out the process.

The DHCP server responds to this request with a DHCP Acknowledge (DHCP ACK) message, which contains the list of configurations previously requested. One of these options (252) contains the location of the PAC file. Finally, the browser can make the download request for this file. The graphical representation of the example is shown in Figure 5.



Figure 5. WPAD via DHCP Source: own work based on [7]

It is important to note that the WPAD via DNS method is the most recommended, since it is supported by most browsers and operating systems. On the other hand, the WPAD via DHCP method demands a greater number of prerequisites, which include: a PAC file, web server, DHCP server, and also for any user, computers must be configured to obtain their network IP address information from the DHCP server. Furthermore, the latter method requires a set of additional steps that must be taken in account by the user for the configuration beforehand; should that not be the case, the DHCP server will not work correctly.

3. Technologies of implementation

Complying with software features, and based on the terms *accuracy*, *reliability* and *robustness* [8], the corresponding components are developed, which improve network end user performance in connecting to the proxy, verifying that the technical attributes defined by Colciencias [9] are met. For the deployment of the project, browser settings and the move into production are taken into account.

a. *Browser configuration*: Automatic proxy settings in the browser are carried out by accessing the

browser settings, followed by the connection configuration. Finally the test URL is entered for automatic proxy configuration

http://127.0.0.1:9645/ or http://127.0.0.1:9646/.

b. *Production*: the packets .deb and .rpm are accessed via the link:

https://github.com/andresvia/udpac/releases/ latest and the corresponding package is installed (.deb, .rpm or binar udpac_linux_amd64) on servers (Debian and Centos):

wpad.udistrital.edu.co PAC URL: http://wpad.udistrital.edu.co:80/

□ wpad.udistritaloas.edu.co

PAC URL:

http://wpad.udistritaloas.edu.co:80/

These packages may be the server itself with two DNS type A registers¹. This server responds with the PAC file on any route, therefore: http://wpad.udistrital.edu.co:80/wpad.dat == http:// wpad.udistrital.edu.co:80/ == http://wpad.udistrital.

edu.co:80/otra_cosa

All routes respond to the PAC file, with the exception of the special metrics that returns information about server performance. For some browsers this is enough. However, to extend coverage you can set the auto-configuration URL through: Policy group (computers in Active Directory) or DHCP settings (computers connected to the network, which automatically configure their IP).

The installed packages contain an application made in GOlang language which has a "Find proxy for" URL function (URL, host) which receives the URL destination and host from which it connects, and within it the following procedures are carried out:

If it isn't possible to establish a connection with the proxy, the failover strategy is used to make a direct connection.

var defaultProxy = "proxy 10.20.4.15:3128; proxy

proxy.udistrital.edu.co:3128; DIRECT";

In this example-case there are subdomains in the institutions that require access to the internet

¹ In this particular case, it was convenient to use the same technique of "split-brain DNS" that the proxy currently has: proxy.udistrital.edu.co

to be solved, so an exception is defined for the Systems Advisory Office (portaloas.udistrital.edu. co), which is served outside the local network. This should be redirected through the proxy as presented in code 1.

Code 1. Applied method of redirection

Likewise, there may be subdomains that do not require access to Internet that must be solved. For example, any request from the browser to domains ending in .udistrital.edu.co or .udistritaloas.edu. co or .local [domain], will not be sent through the proxy. This means that they will connect directly through the local network, providing the customers can resolve these names to current DNS configuration and that there is a network route that can connect them. That is to say, it will act as if the proxy had been manually disabled. This condition is based on the address visited in the address bar and evaluated before any connection or name resolutions occur in the browser; exceptions to this rule must be inserted beforehand. This representation is present in code 2.

> A common domain used for development if (shExpMatch(host, "*.local") || A common domain used as default shExpMatch(host, "*.localdomain") || nip is a clone of XIP shExpMatch(host, "*.nip.io") ||

XIP is a service that maps wildcard domains shExpMatch(host, "*.xip.io") || Official university domain shExpMatch(host, "*.udistrital.edu.co") Non official domain of the systems advisory office shExpMatch(host, "*.udistritaloas.edu.co")) { return "DIRECT"; }

Code 2. Rules applied

One possible scenario: Any request to resolve the DNS name to a local IP (i.e. any private space IP of IPV4) will not be sent through the proxy. Exceptions to this general rule should be inserted before code 3.

if (isPlainHostName(host)) {
 return " DIRECT";
 }
Finally, if none of the above conditions apply, the proxy
 will be used by default
 return defaultProxy;
 }

Code 3. Exceptions applied

Software testing for errors was conducted, seeking not only to identify functional errors, but also to inspect aspects of software quality, maintainability, scalability, efficiency, safety, among others [10]. Figure 6 shows the end user workflow to access the internet according to the Proxy Auto-Configuration via WPAD method.

Another possible scenario: Any requests sent to an unqualified host (e.g. http://www/ or // https:// mail/) will not be sent through the proxy.



Figure 6. Workflow and data within the university network Source: own work

4. Case study and results

In order to observe the impact of the implementation of an auto-configurable proxy in a Higher Education Institution, a university in Latin America was used as case study, with more than 32,000 students, teachers and staff using its Internet services.

For this work, a study of methods and times was conducted in the Faculty of Engineering of the University. The study determined the standard time students spent on manual proxy settings for first time access to the university network through their laptop or smart phone. It then compared the standard time a student spends on Internet access through an access URL, with the time it took once the university proxy servers were auto configured. A statistical formula was used to determine the size of the sample, which took into account that more than 5,500 students registered with the faculty. This figure corresponds to the number of subjects that make up the sample drawn from a population and is necessary for the obtained data to be representative of the population. Formula 1 was then applied.

$$n = \frac{N\sigma^2 Z_{\alpha}^2}{e^2 (N-1) + \sigma^2 Z_{\alpha}^2}$$

Formula 1. Statistical formula was used to determine the size of the sample representative of population

n = size of the sample

N = size of the population. For this study the value used was 5,500 –first to tenth semester– students, taken from different engineering fields

 σ = standard deviation of the population, usually applied when the value is unknown –often a constant value of 0.5 is used, it was the value chosen in this study–

Za: value obtained by confidence levels. This is a constant value that, if unknown, it is based on 95 % confidence, which is equivalent to 1.96 (most common) or based on 99 % confidence, equivalent to 2.58, a value which is left to the criteria of the researcher. For this study, the value of 95 % was chosen.

e = acceptable limit of sampling error; when the value is unknown, it might range from 1 % (0.01) to 9 % (0.09), a choice that is left up to the person

carrying out the poll. For this study the value of 9 % was used.

Given this formula, the recommended sample is 116 students. These were chosen at random and are part of various curriculums within the faculty and are enrolled from first to tenth semester.

This study looked at the time taken by the 116 students who participated. Figure 7 shows the time, in seconds, spent on connecting to the Internet through manual proxy settings. The average time for first time connection to the network through a laptop is 52 seconds, via smartphone: 63 seconds.



Figure 7. Time taken to connect with a manual configuration Source: own work

In a study of conventional methods and duration, the standard time is calculated; that is to say, the "time required by an average skilled operator, working at normal pace, to perform a specified task using a prescribed method" [11]. Formula 2 was then applied [12].

$$T_{std} = T_n * \left(1 + A_{pfd}\right)$$

Formula 2. Method of standard time application Tsdt: standard time

Tn: { $\ Tn: {\ n = } normal time$

Apfd: {\displaystyle A_{\text{pfd}}=} Personal needs, fatigue, and unavoidable delays allowance (pfd): Personal needs, fatigue, and unavoidable delays allowance Given the definition of standard time and considering that the activity does not generate fatigue (it is performed for one minute a day), the value of Personal needs, fatigue, and unavoidable delays allowance (pfd) is almost null for the study, so it is taken as zero. Therefore, the standard connection time through manual proxy settings will be taken from the average value: 52 seconds for a laptop to 63 for a smartphone.

After the implementation of the proposal, a URL connection was granted to the same students participating in the study, in order to calculate the standard connection time. The results are show in Figure 8.





It is possible to observe that the time in seconds spent connecting to the Internet through a proxy auto configuration is significantly less. It can be seen that the average time to connect for the first time to the network through a laptop is 22 seconds and via smartphone 15 seconds. This represents a respective decrease of 58 % and 76 % in time spent.

5. Conclusions and future work

This work proved that this proxy auto configuration for a server in the educational sector is a useful software. It can be used to accelerate access to resources and enhance connection time. Through the study of methods and times, the standard Internet connection time by manual configuration and proxy auto-configuration was determined, for connections via laptop, as well as smartphone. This study evidences the percentage decrease the PAC method gives to end users of a network.

Further work on this project will try to include a web proxy caching based on machine learning, as well as to extend its functionality and analyze students' connection preferences, which could lead to the formulation of hypothesis and recommendations to improve the academic environment within the university network.

References

- C. Jackson, "Method and apparatus for creating proxy auto-configuration file." Google Patents, January 8th, 2004.
- [2] M.Sysel, & O. Dolež, "An Educational HTTP Proxy Server", *Procedia Engineering*, vol. 69, pp. 128-132, 2014. doi:10.1016/j.proeng.2014.02.212
- [3] L. L. C. Books & G. B. LLC, Proxy Servers: Proxy Server, Wingate, Tor, Proxomitron, Proxy Auto-Config, Java Anon Proxy, Sun Java System Web Proxy Server, Web Cache. Memphis, Tennessee: General Books LLC, 2010.
- [4] S. Shiwani, S. Kumar, V. Chandra & S. Bansal, "Performance Measurements: Proxy Server For Various Operating Systems", en 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, India, Nov. 27-29, 2014, pp. 98-102.
- [5] Internet Protocol, Defense Advanced Research Projects Agency, Darpa Standard 791, 1981. Available: https://tools.ietf.org/html/rfc791
- [6] Transmission Control Protocol, Defense Advanced Research Projects Agency, Darpa Standard 793, 1981. Available: https://www.ietf.org/rfc/rfc793.txt
- [7] P. Hayes. (2007). Guiding principle behind Find-ProxyForURL, Find and proxy for URL iniciative [Online]. Available: http://findproxyforurl.com/
- [8] A. Tucker, *Computer science hand book*, 2nd ed. Boca Ratón, Florida: crc Press, 2004, p. 2752.
- [9] Colciencias. (2004, March, 19). Resolución 00285. Por la cual se establecen los criterios para la asignación de puntajes por productividad académica en producción de software, para los docentes de entidades públicas o estatales. Available: https://www. uniatlantico.edu.co/uatlantico/sites/default/files/ docencia/pdf/Resolucion%20285%20Reglamentacion%20Software.pdf

- [10] I. Sommerville, Software Engineering. 9th ed. United States: Pearson Education, 2011, p. 773.
- [11] M. P. Groover, *Work Systems and the Methods, Measurement, and Management of Work.* United States: Pearson Prentice Hall, 2007, p. 792.
- [12] A. Freivalds & B. Niebel. (2013). Niebel's Methods, Standards, & Work Design. 13th ed.United States: Mc Graw Hill, p. 752.